

Effect of analysis strategy on reproducibility of detected fMRI activations

E. Salli^{1,2}, H.J. Aronen², A. Korvenoja^{2,3}, A. Visa⁴, and T. Katila^{1,3}

¹Laboratory of Biomedical Engineering, Helsinki University of Technology, FIN-02015 HUT, Finland;

²Department of Radiology, ³BioMag Laboratory, Helsinki University Central Hospital, FIN-00029 HUS, Finland; ⁴Signal Processing Laboratory, Tampere University of Technology, FIN-33101 Tampere, Finland

1 Introduction

The primary goal of image analysis in functional magnetic resonance imaging (fMRI) activation studies is usually to detect and delineate the image areas that have a signal intensity time course, which can be related to the experimental parameters. The task is challenging because the images are noisy and often corrupted by motion. The problem is typically solved using a statistical testing procedure. First, a statistical parametric map (SPM) is created. Thereafter, either a non-active or active state is assigned to each voxel. We will call this the segmentation phase, as the goal of this step is to divide the statistical parametric map into non-active and active regions.

The statistical parametric maps are often computed using the general linear model that subsume for example the simple t test [1, 2]. A commonly used segmentation method is intensity thresholding. The voxels whose statistical value is larger than a predefined threshold, directly related to the significance level of a statistical test, are classified as active.

One way to improve the segmentation is to utilize an assumption that the probability of the activation is related to the existence of activation in the spatial neighborhood of a voxel. This *a priori* assumption can be incorporated into the classification procedure by using so called contextual methods.

Previously, a computationally efficient contextual clustering algorithm for the segmentation of statistical parametric maps was introduced [3]. In the contextual clustering algorithm, both statistical parametric values and classification information from the neighborhood of each voxel are iteratively used to make a decision whether a voxel is active or not.

The goal of the present study was to analyze the reproducibility of segmentation. Especially, the effect of contextual information on the reproducibility was studied. Additionally, we studied if a reproducibility index could be used to choose the segmentation parameters in an objective way.

2 Materials and Methods

2.1 Subjects and data acquisition

MR imaging was performed with a Siemens Vision 1.5 T MRI scanner (Siemens AG, Erlangen, Germany) at the Department of Radiology, Helsinki University Central Hospital.

Functional MR images were acquired using a gradient-echo echo-planar imaging sequence (EPI) (TE 70 ms, TR 2.083 s, flip angle 90°, field of view 256 mm, matrix 64 × 64, 16 slices, slice thickness 3.0 mm, gap 1.0 mm). Four motor experiment studies were performed on a right-handed healthy volunteer. To minimize head movement, a head-supporting vacuum cast was used. The magnetic field was globally shimmed prior to imaging.

In the experiments, volunteer was instructed to flex his right wrist during the presence of the character + on a screen and rest during the presence of the character x. The paradigm consisted of four rest and four motor execution blocks, lasting 15 scans each. In addition, prior to the first rest epoch of each study, 8 scans were acquired to allow the MRI to reach a steady state of longitudinal magnetization.

2.2 Preprocessing and computation of maps

Statistical parametric maps were computed using the SPM99 software (Wellcome Department of Cognitive Neurology, London, UK) following the guidelines for a basic statistical analysis. First, the images were realigned in order to remove movement-related variance components. Sinc interpolation was used in the transformation. The statistical parametric maps (t maps) were computed by using the general linear model. The linear model for the signal was specified to be the fixed box-car function convolved with model hemodynamic response function (hrf). Estimated movement parameters were used as covariates in the analysis. Serial correlations were dealt with temporal filtering of the data. Cut-off period of the high-pass filter was set to 125 s. Temporal low-pass filtering was done to allow for a proper assessment

of degrees of freedom. Finally, the t maps were converted to z maps.

2.3 Segmentation of z maps

We follow here the notation in which the activations have a positive mean. In the thresholding, a voxel i is classified as active if

$$z_i > T \quad (1)$$

where z_i is the z value of a voxel i and T is a pre-defined intensity threshold. Otherwise, the voxel is treated as non-activated.

The contextual clustering rule used in this paper is as follows [3]: First, as an initialization step the classification is initialized using Eq. (1). After the initialization, the voxels are re-classified. In the re-classification, a voxel is considered as active if

$$z_i + \frac{\beta}{T}(u_i - N_n/2) > T \quad (2)$$

and otherwise as non-active. Constant N_n is the number of neighbors under consideration. By defining the neighborhood to consist of 26 closest voxels, $N_n = 26$. Variable u_i is the number of currently active neighborhood voxels of voxel i . The parameter β determines the weighting of neighborhood information, and when positive, encourages neighbors to be of like class. One way to set the β is to write

$$\beta = \frac{T^2}{s}. \quad (3)$$

Parameter s is a user specified parameter and may have any real positive value. Intuitively, s can be understood as a required excess of activated voxels ($= u_i - N_n/2$) in the neighborhood of voxel i to transform a value of zero to the level of T . Classification rule (2) is repeated until convergence or oscillation between states occurs. During each iteration the classification and u_i 's are updated. It is easy to see that as s approaches ∞ the contextual clustering approaches the voxel-by-voxel thresholding technique. As s approaches zero, the contextual clustering approaches a recursive majority vote classification.

The data sets were contextually segmented with parameter values $T = 0.1, 0.2 \dots 7.0$ and $s = 0.5, 2, 6, 10, 20, 50$. In addition, the segmentation was performed using the voxel-by-voxel thresholding.

2.4 Reproducibility

Reproducibility of the segmentation was studied by comparing the analyzed images of four motor activation studies. Segmented (i.e. binary) maps of all four studies were summed up voxel by voxel. Voxel value of the sum images, R_i , represents the number of studies (from zero to four) in which the voxel i is classified as active. This kind of a sum image has been named as a reliability map [4] and we will adopt the term here. The reliability map gives an idea about the reproducibility of the method. If all the non-zero voxels have maximum value (in this case the value of four) the results have been perfectly reproducible. If all the non-zero voxels have a value of one the results have not been reproducible at all.

In order to quantitatively compare the effect of segmentation parameters on the reproducibility a measure of the reproducibility, R_m , was defined. We defined the reproducibility index R_m to be the mean value of all non-zero values of R_i .

The reliability maps were transformed to the coordinates of the anatomical MR image sets for the visualization purposes using slice-positioning information available from the image file headers.

3 Results

Figure 1 shows the reproducibility index achieved with different segmentation parameters. It is seen that as s increases, the contextual clustering approaches voxel-by-voxel thresholding technique. This is expected result because the contextual clustering approaches the thresholding when $s \rightarrow \infty$. Most of the curves have two local maxima. The first maximum is obtained with low T values. It can be explained easily by the fact that with small T values the probability of false activation is so high that it is likely that a voxel is classified falsely as active also in other experiments. For example, if thresholding were done using very small T ($T \rightarrow -\infty$), then all voxels would be classified as active and the reproducibility index would be $R_m = 4$. Hence, the first maximum is not interesting. At the second maximum, the probability of false activations is small and the detected activations are likely caused by true activations. This latter maximum is the interesting one.

The highest reproducibility index $R_m = 3.1$ was achieved with parameter combination $s = 2$, $T = 1.4$. As s was increased, the maximum of reproducibility indexes over the T values decreased, being

$R_m = 2.4$ for the thresholding ($T = 5.1$).

Figure 2 shows two reliability maps superimposed on the anatomical MR images. The value R_i in the reliability maps is the number of studies (from 0 to 4) in which the voxel was classified as active.

It should be noted that the images shown in Fig. 2 do not have same false-positive rates, or significance levels. T values were chosen so that R_m was maximized with given s . A simple simulation experiment was performed to assess the statistical significance of the results. 50000 simulated statistical parametric maps ($64 \times 64 \times 16$ voxels in size) without activations were created and segmented. With thresholding ($T = 5.1$) at least one (falsely) active voxel was detected in 542 images (545 active voxels in total). With contextual clustering using values $s = 20, T = 3.1$ an activation was detected in 499 images (502 active voxels in total). With $s = 2, T = 1.4$ no activations were detected. This result indicates that the activations of Fig. 2(b) were obtained by using much stricter criterion than in the Fig. 2(a) or Fig. 2(c). However, the results of the simulations should be seen only as rough approximations as the simulated data were sampled from a pseudo-random standard normal distribution and the voxels were spatially uncorrelated.

Activations of primary sensorimotor cortex were detected in all cases of Fig. 2. Larger number of $R_i = 4$ voxels (voxels active in all 4 experiments) were detected when contextual information was utilized. Activations of the supplementary motor area (SMA) were not detected with $s = 2$ (Fig. 2(b)). Apparent explanation is the strict statistical criterion.

3.1 Discussion

We have studied the reproducibility of the fMRI activation patterns. Especially, the effect of the contextual information was studied. If functional MRI is used as a physiological constraint for electromagnetic inverse problem, it is of great importance that activations areas are delineated reliably [5]. Our view is that the higher reproducibility of the fMRI localization results increases the possibility of using fMRI as a constraint in MEG.

There are multiple sources of variation that limit the possibilities to obtain reproducible results. The thermal noise follows the normal distribution well and exists independently in neighboring voxels. Therefore the effects of this noise source can be handled statistically. The movements are significant sources of variation. In this study, we used head supporting vac-

uum cast and realignment algorithm in order to reduce the effect of these variation sources. Magnetic field instabilities [6] and physiological effects may cause drifts and temporal or spatial correlations to the data. Therefore we filtered the data in temporal domain and used reduced degrees of freedoms. Some variation may arise from the timing errors between the stimulus signal, imaging sequence and actual movement execution of a hand. The variation caused by these timing errors is believed to be negligible.

The contextual clustering method uses as an input only the z map (e.g. $64 \times 64 \times 16$) instead of the whole data. Hence, the time required to run the contextual clustering algorithm is only a few seconds with a 500 MHz Pentium III.

In this short paper we have not studied the effect of segmentation parameters to the false positive rates, sensitivity or segmentation accuracy. These issues were studied in [3]. In the statistical testing techniques, the parameter values (e.g. thresholds) are chosen so that a desired false-positive rate is achieved. There are a few problems associated with the testing approach. The first one is that the selection of accepted false-positive rate is somewhat arbitrarily. The second problem is that the accurate determination of decision parameter values is difficult due to spatial autocorrelations and other non-idealities in the data. Reasonable results were obtained when the parameters were chosen so that the defined reproducibility index was maximized. This suggests that it might be possible to use the reproducibility index as an objective criteria on to choose the parameter values for a segmentation algorithm. The obvious drawback of the approach is that in order to compute a reproducibility index an experiment must be repeated or one experiment divided into parts. In addition, reasonable results are expected only if there are truly active regions in the data that can be detected. Hence, the reproducibility index used as a method to choose the decision parameter can be seen as a complementary activation visualization method after the existence of activations has been statistically reasoned.

The main result this study was that the use of spatial neighborhood information increases reproducibility. According to the results, highest reproducibility indices were achieved when high weight was given to contextual information ($s = 2$). However, in order to preserve small activation regions and better segmentation accuracy [3] it may be useful to use somewhat smaller weights.

Acknowledgements

This work was supported by the Foundation of Technology in Finland, Jenny and Antti Wihuri Foundation, Radiological Society of Finland, Academy of Finland, Sigrid Jusélius Foundation and Helsinki University Central Hospital Research Grants.

References

- [1] K. J. Friston, A. P. Holmes, K. J. Worsley, J. B. Poline, C. D. Frith, and R. S. J. Frackowiak. Statistical parametric maps in functional imaging: a general linear approach. *Human Brain Mapping*, 2:189–210, 1995.
- [2] K. J. Friston, A. P. Holmes, J. B. Poline, P. J. Grasby, S. C. Williams, R. S. Frackowiak, and R. Turner. Analysis of fMRI time-series revisited. *NeuroImage*, 2:45–53, 1995.
- [3] E. Salli, A. Visa, H. J. Aronen, A. Korvenoja, and T. Katila. Statistical segmentation of fMRI activations using contextual clustering. *Proc. of MICCAI'99. Lecture Notes in Computer Science*, 1679:481–488, 1999.
- [4] C. R. Genovese, D. C. Noll, and W. F. Eddy. Estimating test-retest reliability in functional MR imaging I: Statistical methodology. *Magnetic Resonance in Medicine*, 38:497–507, 1997.
- [5] A. Korvenoja, J. Huttunen, E. Salli, H. Pohjonen, S. Martinkauppi, J. M. Palva, L. Lauronen, J. Virtanen, R. J. Ilmoniemi, and H. J. Aronen. Activation of multiple cortical areas in response to somatosensory stimulation: Combined magnetoencephalographic and functional magnetic resonance imaging. *Human Brain Mapping*, 8:13–27, 1999.
- [6] A. M. Smith, B. K. Lewis, U. E. Ruttimann, F. Q. Ye, T. M. Sinnwell, Y. Yang, J. H. Duyn, and J. A. Frank. Investigation of low frequency drift in fMRI signal. *NeuroImage*, 9:526–533, 1999.

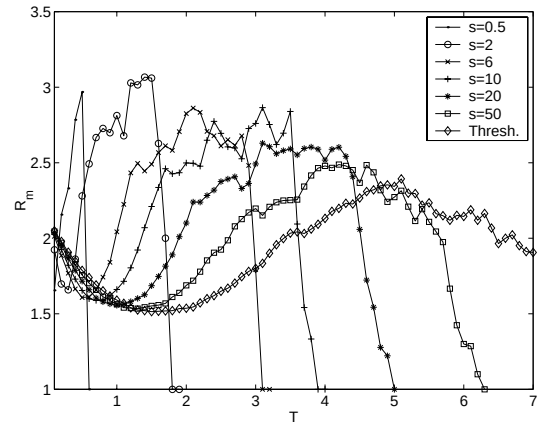


Figure 1: Reproducibility index, R_m , obtained with varied parameter values s and T . R_m is defined as a mean of non-zero R_i values. The R_i value of a voxel is the number of studies (from 0 to 4) in which the voxel is classified as active (see Fig. 2).

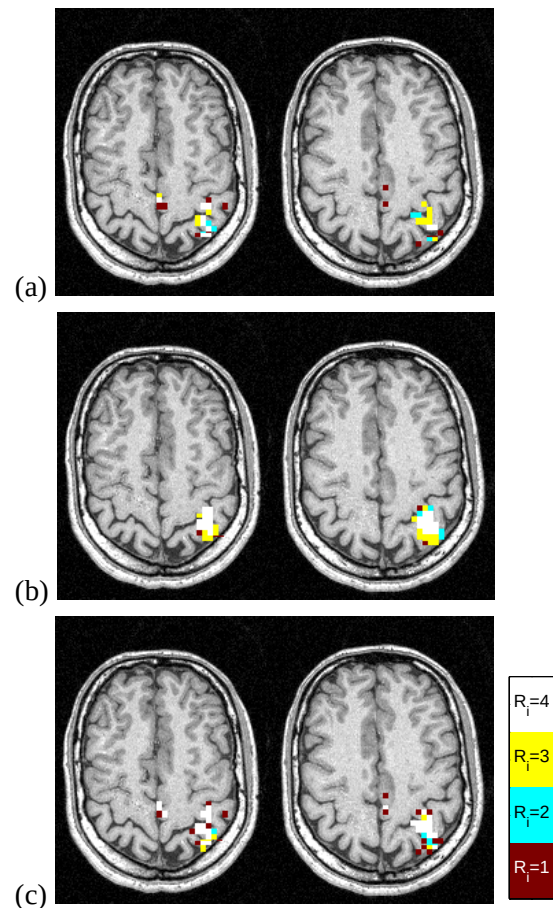


Figure 2: Reliability maps. The colored voxels represent the number of studies (R_i) in which the voxel was classified as active. (a) Voxel-by-voxel thresholding $T = 5.1$, (b) Contextual clustering $s = 2, T = 1.4$, (c) $s = 20, T = 3.1$. With contextual information the proportion of $R_i = 4$ voxels increased.